

作业五：机器翻译

本实验旨在构建一个基于注意力机制的序列到序列模型，实现从法语到英语的机器翻译任务。模型采用编码器-解码器架构，编码器与解码器均为 GRU 网络，并引入注意力机制以提升翻译质量。实验基于法语-英语平行语料库，包含约 13.5 万句对，经预处理得到有效句对 95170 个。我们通过对比不同超参数配置下的模型表现，分析学习率、隐藏层大小和批大小对翻译性能的影响，最终获得较优模型配置。

一、实验目的与模型架构

1.1 实验目的

- 实现基于注意力机制的序列到序列翻译模型
- 探究关键超参数对模型性能的影响
- 通过实验优化模型，提高法语到英语的翻译准确率

1.2 模型架构

模型采用编码器-解码器结构：

- 编码器：单向 GRU 网络，将输入的法语句子编码为隐藏状态序列
- 解码器：基于注意力机制的 GRU 网络，依据编码器输出及当前隐藏状态生成对应英语单词
- 注意力机制：帮助解码器在生成每个词时动态关注输入序列的相关部分

1.3 数据集与预处理

- 使用法语-英语平行语料库，初始包含约 13.5 万句对
- 经清洗与过滤，保留 95170 个有效句对
- 法语词典大小：7874 词，英语词典大小：5506 词
- 按 9:1 划分训练集与测试集

1.4 训练设置

- 损失函数：交叉熵损失
- 优化器：Adam
- 基准参数：学习率 0.001，批大小 64，隐藏层维度 512，迭代次数 10000

二、实验结果

在基准参数配置下，模型在测试集上达到 25.376% 的准确率，符合课程基本要求。模型能够对部分简单句子实现正确翻译，例如：

法语输入：tom ravalà sa fierté.

英语输出：tom swallowed his pride . <EOS>

但面对稍复杂的句子时，模型仍出现语义缺失或错误，例如：

法语输入：grandis un peu .

英语输出：a little bit . <EOS> (缺失主要动词"grow up")

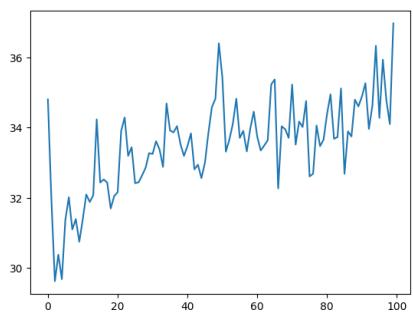
训练过程：

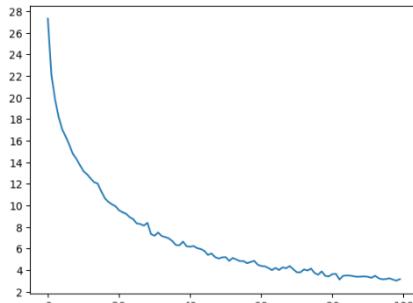
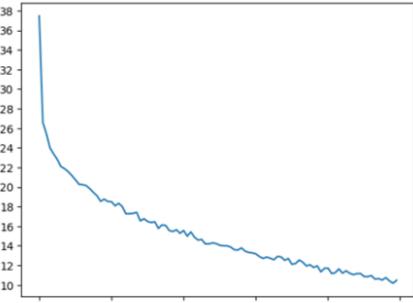
- 损失值：3.328
- 训练时间：约 10 分钟
- 损失曲线平稳下降，呈现良好收敛趋势

三、超参数调优实验

为研究超参数的影响，分别调整学习率、隐藏层大小和批大小进行对照实验。

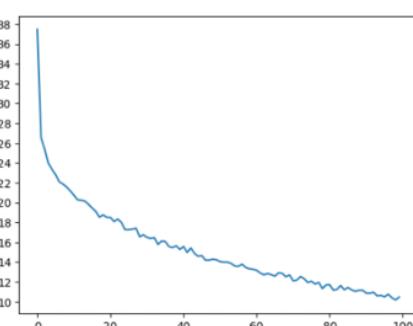
3.1 学习率的影响

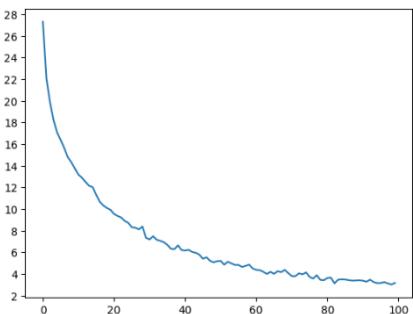
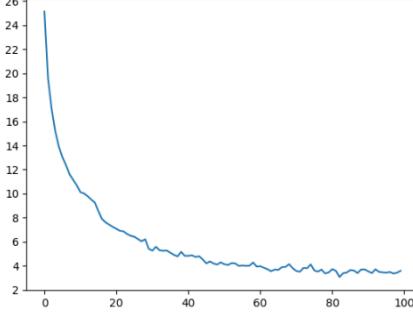
| 学习率 | 准确率 (%) | 损失函数曲线图 | 损失函数曲线图特点 |
|------|---------|--|-----------|
| 0.01 | 0.315 |  | 波动剧烈，未收敛 |

| | | | |
|--------|--------|--|-----------|
| 0.001 | 25.376 |  | 平稳下降，良好收敛 |
| 0.0001 | 6.741 |  | 下降缓慢，收敛不足 |

结论：学习率过高（0.01）导致梯度震荡，无法收敛；学习率过低（0.0001）则收敛缓慢，模型未能充分学习；0.001 的学习率表现最佳。

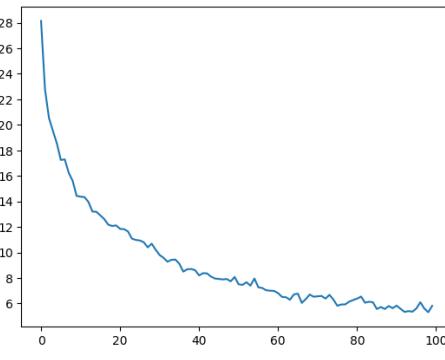
3.2 隐藏层大小的影响

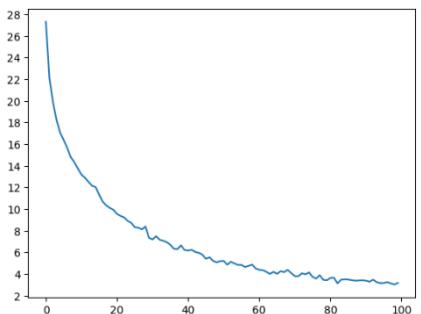
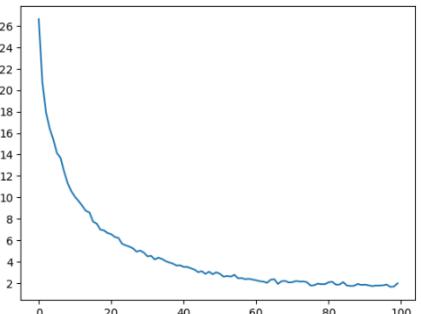
| 隐藏层大小 | 准确率 (%) | 损失函数曲线图 | 损失函数曲线图特点 |
|-------|---------|--|-----------|
| 256 | 21.748 |  | 平稳下降 |

| | | | |
|------|--------|--|------|
| 512 | 25.376 |  | 平稳下降 |
| 1024 | 25.929 |  | 平稳下降 |

结论：隐藏层增大可提升模型表达能力，改善对复杂语义的捕捉能力，但也会轻微增加训练时间。

3.3 批大小的影响

| 批大小 | 准确率 (%) | 损失函数曲线图 | 损失函数曲线图特点 |
|-----|---------|--|-----------|
| 32 | 21.141 |  | 平稳下降 |

| | | | |
|-----|--------|--|------------|
| 64 | 25.376 |  | 平稳下降 |
| 128 | 27.267 |  | 平稳下降, 收敛最佳 |

结论：批大小增大有助于获得更稳定的梯度估计，提升模型泛化性能，但训练时间相应增加。

四、翻译结果示例分析

以法语句子 "Tu es un enfant délicieux" 为例，不同实验的输出如下：

| 实验配置 | 模型输出 | 评价 |
|------------------------------|----------------------------------|------------|
| 基准 (lr=0.001, hs=512, bs=64) | you re a kid child . <EOS> | 语义偏离 |
| 学习率 0.01 | you re a <SOS> <SOS>... | 无法生成有效句子 |
| 学习率 0.0001 | you re a a liar . <EOS> | 语义错误 |
| 隐藏层 256 | you re a strong girl . <EOS> | 部分正确，但性别错误 |
| 隐藏层 1024 | you are a terrible child . <EOS> | 语义错误 |
| 批大小 32 | you re a kid kid . <EOS> | 基本正确，但有重复词 |
| 批大小 128 | you re a wonderful boy . <EOS> | 较为接近原意 |

五、结论

通过本次实验，我们看到学习率对模型收敛具有关键影响，过高或过低均会导致性能下降。隐藏层大小也会影响模型表达能力，增大隐藏层可提升翻译质量，但也会增加计算开销。最后，批大小需要在训练效率与梯度稳定性之间权衡，适当增大批大小有助于提升模型泛化能力。